

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version:

Valdez Banda, Osiris; Goerlandt, Floris; Salokannel, Johanna; van Gelder, Pieter H.A.J.M.

An initial evaluation framework for the design and operational use of maritime STAMP-based safety management systems

Published in: WMU JOURNAL OF MARITIME AFFAIRS

DOI: 10.1007/s13437-019-00180-0

Published: 01/01/2019

Document Version Publisher's PDF, also known as Version of record

Please cite the original version: Valdez Banda, O., Goerlandt, F., Salokannel, J., & van Gelder, P. H. A. J. M. (2019). An initial evaluation framework for the design and operational use of maritime STAMP-based safety management systems. WMU JOURNAL OF MARITIME AFFAIRS. <https://doi.org/10.1007/s13437-019-00180-0>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



An initial evaluation framework for the design and operational use of maritime STAMP-based safety management systems

Osiris A. Valdez Banda, et al. *[full author details at the end of the article]*

Received: 10 July 2019 / Accepted: 24 September 2019/Published online: 30 October 2019
© The Author(s) 2019

Abstract

A safety management system (SMS) is the common means used by organizations to assess organizational performance with respect to the safety and well-being of people, property and the natural ecosystem. A SMS provides confidence to diverse stakeholders that organizational safety is at an appropriate level and fulfils the applicable regulatory standards. As a multifaceted system for organizational safety assessment, ensurance and assurance, the evaluation of the design and operational use of SMS is a complex process. An evaluation needs to provide evidence about how well the design and operation of an SMS complies with applicable standards and how well the methods used in the SMS implementation support the organizational policies and practical work. In the maritime domain, SMS is broadly applied. However, there are few theoretically rooted SMS design approaches, and there is a lack of frameworks to evaluate how well the SMS is designed and how effectively it operates. This paper proposes an initial evaluation framework for the design and operational use of a maritime SMS design approach based on Systems-Theoretic Accident Model and Processes (STAMP), realist evaluation and Bayesian Networks. This framework is applied for a case study of vessel traffic services (VTS) Finland to test its relevance and ability to guide the SMS design. The experiences gained in the case study, and the related discussion on the framework, can guide further research in this area. Ultimately, the work can be used as a basis for developing maritime SMS auditing processes, based on specific theoretical and methodological approaches.

Keywords Evaluation framework · Safety management system · Systems-Theoretic Accident Model and Processes (STAMP) · Maritime safety · Vessel traffic services

Highlights:

- A framework to evaluate the design and operation of safety management systems (SMS)
- The framework detects key aspects of the design and operation of maritime SMS
- The framework focuses on evaluating the theories, practices and standards of the SMS
- The framework evaluates the design of a SMS for vessel traffic services in Finland

1 Introduction

A safety management system (SMS) is the common organizational vehicle to assess and assure the safety of people, property and the natural ecosystem (Reiman and Rollenhagen 2011, 2014). By definition, an SMS is a system designed to develop, plan, measure, analyse and control the overall safety performance of an organization and to guide safety management through selecting appropriate safety ensurance activities (Hale et al. 1997; Hollnagel et al. 2008; Valdez Banda et al. 2016). In this effort, the SMS must comply with the requirements of applicable safety regulations, which is necessary for safety assurance, confirmation of compliance and certification (Kelly 2017).

Safety management systems appear in various forms, corresponding to different historical development paths in safety science. These differences are articulated through adherence to different conceptualizations of what safety is, alternative accident and organizational theories, and different methods and techniques underlying SMS development and application (Li and Guldenmund 2018). Furthermore, given the importance of SMS certification for organizations—such as for legal reasons—SMS design and application also differ between application domains due to differences between the regulatory regimes of various industries (Maurino 2017). Hence, differences between applicable standards affect how exactly safety management systems are implemented in different organizations, and how audits or regulatory compliance are performed (Li and Guldenmund 2018).

Nevertheless, attempts have been made to define a generic structure of safety management systems; see e.g. Hale (2005), who argues that all SMS consist of a risk control system and a learning system, each consisting of several sub-elements. According to Thomas (2012) and Maurino (2017), essential components of any SMS across regulatory and application domains include the following: (i) identification of safety hazards, (ii) remedial action to maintain safety performance, (iii) continuous monitoring and regular assessment of safety performance and (iv) continuous improvement of the overall performance of the SMS, see also e.g. ICAO (2009) and IMO (1993). Fernández-Muñiz et al. (2007) found that the key dimensions of a safety management system across industries consist of the following: (i) the safety policy, which includes the organization's commitment to safety, formalizing principles, objectives, strategies and guidelines; (ii) incentives for employee participation, aimed at promoting safe behaviour and personnel involvement in decision processes; (iii) training and development of employee competences; (iv) communication and information transfer about risks and risk controls; (v) planning, addressing both prevention and emergency response and (vi) control and review activities.

Safety management systems essentially take a business management approach to safety (Thomas 2012; Maurino 2017), and even though this bureaucratization of safety work may have benefits and serve specific functions in organizations (Rae and Provan 2018), it can also have secondary effects involving accountability, reduced marginal yield of safety initiatives and stifling of organizational freedom and innovation (Dekker 2014). Furthermore, there are often significant practical challenges to implementing SMS in organizations (Gerde 2015; Lappalainen et al. 2014).

From a scientific perspective, an important issue is whether increased proceduralization of safety, such as through implementation of SMS, actually enhances safety, and if so, how this relates to other positive or negative effects on organizational performance (Bieder and

Bourrier 2013). This question of scientific evidence for safety practices is a more general concern in safety science (Goerlandt et al. 2017; Hale 2014; Le Coze et al. 2014). However, focusing on SMS, this is a complicated issue due to the wide range of aspects influencing the design and application of SMS (Li and Guldenmund 2018) and the variations in specific, practical implementations (Thomas 2012).

There is little systematic evidence concerning the validity or effectiveness of SMS. It is plausible that SMS indeed enhances safety performance and outcomes. While there is some supporting empirical evidence to this effect (LaMontagne et al. 2004), much of this evidence is of comparatively low quality due to, for instance, methodological problems with common method variance, and there is some ambiguity between the results of different studies (Thomas 2012). Especially safety outcomes related to low-probability, high-consequence activities are practically impossible to empirically ascertain directly in a statistical sense due to the low rate of process accidents (Rae 2018).

In the maritime domain, SMS has been an object of research primarily in the context of the implementation of the International Safety Management (ISM) Code, which is seen as an important vehicle to focus on human and organizational factors in the development of maritime safety (Schröder-Hinrichs 2010). Previous research has focused on how shipping accidents and incidents relate to functional sections of the ISM Code (Batalden and Sydnese 2014); how the ISM code is applied in shipping practice (Lappalainen et al. 2011; Batalden and Sydnese 2015); what are the relationships between regulation, safety culture and safety management (Kongsvik et al. 2014) and whether the employment and social conditions in maritime shipping align with the need for self-regulation necessary to successfully implement the ISM Code (Bhattacharya 2012).

Research has also been undertaken to assess the extent to which the ISM Code has contributed to improving safety, in line with the questions on the usefulness of SMS in the general safety science literature, as outlined above. Work addressing this important question has been reported by Tzannatos and Kokotos (2009), Oltedal (2010), Pantouvakis and Karakasni (2016) and Karakasni et al. (2018). While the results are not univocal, it appears that overall the ISM Code has had a positive effect on safety performance in shipping.

In other application domains, such as the chemical process industry (Basso et al. 2004), construction (Teo et al. 2006) and aviation (Chang et al. 2015), several methods have been proposed for evaluating the performance of safety management systems. To the best of the authors' knowledge, no comparable methods have been proposed in the maritime domain. The work by Celik (2009) has some similarities, but this focuses on the extent to which an SMS based on the ISM Code is implemented based on the principles of the ISO9001 quality standard, rather than on evaluating the SMS in regard to the safety assessment, ensurance and assurance functions it aims to serve.

A relatively recent area of academic attention is the development of approaches to design a maritime SMS. Akyuz and Celik (2014) propose a method to identify and prioritize key performance indicators (KPIs) for designing a maritime SMS based on the analytical hierarchy process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Valdez Banda et al. (2016) propose a method for extracting KPIs from maritime safety management standards, based on realist evaluation, giving an alternative lens to empiricist evaluation techniques, and expert elicitation. An important issue in safety science is the need to distinguish occupational safety

incidents and organizational accidents due to the different causal factors involved in their occurrence (Meyer and Reniers 2016; Størkersen et al. 2017). Furthermore, the co-existence of different organizational accident theories as described by Qureshi (2007) affects the design of the specific components of an SMS and the definition of the KPIs used to assess organizational performance (Lofquist 2017; Batalden and Oltedal 2018). Hence, acknowledging the importance of rooting the development of an SMS in a well-defined safety and accident theory basis, Valdez Banda and Goerlandt (2018) propose a design approach for maritime SMS based on the Systems-Theoretic Accident Model and Processes (STAMP). This approach applies the Systems-Theoretic Process Analysis (STPA) for identifying and analysing hazards and safety controls, realist evaluation to identify KPIs and Bayesian Networks as a measurement and decision support tool.

Considering the lack of approaches to evaluate the design and operational use of maritime SMS, especially where these are based on design approaches rooted in specific accident theories, this article aims to propose an evaluation framework for the design and operational use of SMS based on a specific design approach, as presented in Valdez Banda and Goerlandt (2018). This evaluation framework is applied to a case study for the design of an SMS for vessel traffic services (VTS) Finland, providing VTSs to merchant shipping and other marine traffic and maintaining safety radio operations, which is elaborated in Valdez Banda and Goerlandt (2017, 2018). As no earlier work in this area has been undertaken, the framework is proposed as an initial approach for evaluation of maritime SMS design and operational use. The results of the case study and insights gained are used to formulate future research directions, contributing to the development of a theoretically founded maritime SMS auditing process.

The remainder of this article is organized as follows. Section 2 presents the background, scope and research objectives. Section 3 introduces the proposed framework for evaluating safety management systems. Section 4 presents a case study where the evaluation framework is applied to an SMS designed for VTS Finland. Section 5 presents the results of this application. Section 6 discusses the study results and Section 7 presents the study conclusions.

2 Background, scope and research objectives

2.1 Safety management systems: generic model of related issues

In a recent review article, Li and Guldenmund (2018) present a schematic model that provides an overview of general issues related to SMS. To facilitate the work presented in the remainder of this article, this model is shown in Fig. 1 and briefly outlined. Three levels are distinguished: the theoretical level, the practical level and the standards level. Within these levels, theories, methods/techniques, audit tools and standards serve specific functions in the design, application, auditing and compliance assurance of SMS.

At the theoretical level, the focus is on the theoretical models underlying the design, operational use and auditing of an SMS. This includes how safety is understood as a concept, which accident theories are adhered to, i.e. how accidents are considered to occur and how they can be prevented; how hazard control is managed, i.e. through what organizational activities and how these aspects are logically linked. The standards level concerns the regulatory requirements, focusing on how standards and/or

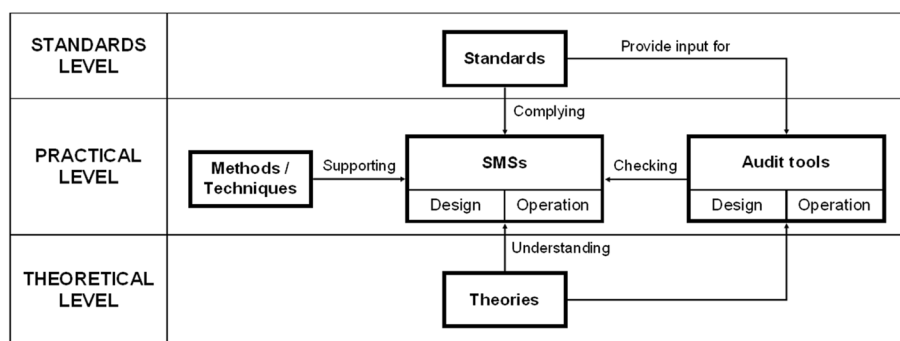


Fig. 1 Generic model of issues related to safety management systems, adapted from Li and Guldenmund (2018)

guidelines shape the design, use and auditing of SMS. The practical level is where the actual design, operational use and auditing of an SMS take place, receiving inputs from both the theoretical level and the standards level. This level includes what methods and techniques are used to analyse the hazards and identify control actions; how key performance indicators (KPIs) are selected, defined and measured; how these KPIs are linked to control actions; what tools are used to connect the SMS to organizational activities and what tools and processes are used for auditing the system in the context of improving SMS performance and for regulatory compliance monitoring.

2.2 Evaluation and validation

As outlined in the introduction, while evidence suggests that SMS contribute to organizational safety, there is relatively little empirical evidence about which specific components or processes are involved in this. This is a more general cause of concern in safety science: while there are many co-existing paradigms and corresponding approaches to measure, assess, ensure and assure safety (Rae 2015), comparatively little work has been carried out to empirically ascertain their validity (Möller et al. 2018).

Acknowledging the importance of clarity about the intended meaning of key concepts in developing safety science (Aven 2014), this section briefly specifies what is meant in this paper with the concepts of evaluation and validation in the context of SMS and organizational safety performance. The explicit purpose of these definitions is to serve stipulative and distinguishing functions; that is, they aim to clarify how these terms are to be understood in the bounded context of the proposed framework, and how they differ from one another (Rosa 2003). These definitions aim to clarify the exact scope of the proposed framework in Section 3 and the scope of the case study in Section 4.

Evaluation is understood as the assessment of the extent to which the elements considered in the SMS design are deemed appropriate for controlling the hazards to which the system is exposed; whether those elements are appropriately linked to measure, analyse and control system safety and whether the measures (e.g. indicators) used are appropriate. It also concerns whether the users of the SMS find that it is well integrated into their work practices, using appropriate methods and tools, and that it serves its purpose effectively in safety assessment, safety-related decision-making and safety assurance. Considering the model in Fig. 1, evaluation as a concept is therefore closely linked to

audit tools, having the explicit aim of detecting both positive and negative aspects of the SMS design and operational use, based on which improvements can be made.

Validation is understood as a process for establishing pragmatic validity, which is the condition where the SMS successfully achieves what it aims to achieve, in line with terminology suggested by Rae et al. (2014) and Goerlandt et al. (2017). Validation thus focuses on whether the implementation of the SMS actually leads to improved safety performance through better safety outcomes, which is typically stated as the main reason for adopting SMS in organizational practices (Thomas 2012).

2.3 Specific SMS design approach and case study in focus

The evaluation framework proposed in this study is applied to evaluate the design of an SMS proposed to VTS Finland, based on an approach introduced in Valdez Banda and Goerlandt (2018). The process of designing this SMS covers the three levels of related issues of safety management systems proposed in Li and Guldenmund (2018).

At the theoretical level, safety is defined as a system property and therefore, it should be managed at the system level (Leveson 2011). Systems-Theoretic Accident Model and Processes (STAMP) is taken as a model for organizational accidents, based on which the SMS is structured into six hierarchical levels. These levels aid in the systemization of the reasons for design decisions in constructing the elements in organizational safety management and the interactions between these. The theoretical level is complemented with proposed KPIs, which are used in a Plan-Do-Check-Act (PDCA) cycle to guide decision-making and safety assurance actions, conditional to different levels of organizational safety performance as measured by the SMS (Chang and Liang 2009).

At the practical level, the design process executes the Systems-Theoretic Process Analysis (STPA), a hazard analysis method based on STAMP, which includes the definition of accident scenarios covering design errors, component interactions and other social, organizational and management factors in the analysis (Leveson 2011). This states the purpose and scope of the system and defines the safety controls that are represented in the SMS. In addition, KPIs for the SMS are defined by applying the realist evaluation method, proposed by Pawson and Tilley (1997). These KPIs are integrated into a Bayesian Network (BN) model that implements the complete PDCA process. This BN model serves as a practical tool to operationalize the SMS and aims to support the planning, execution, review and improvement of SMS performance.

At the standards level, the process to design the SMS for VTS Finland includes a detailed review of standardized work procedures and directions for regulatory compliance. The aim is to establish efficient integration between regulatory demands and the actual practices in the organization. This task includes a review of the VTS quality management system and its approach to fulfilling the regulatory demands of the International Association of Lighthouse Authorities (IALA). The development of the SMS design for VTS Finland is elaborated in Valdez Banda and Goerlandt (2017, 2018).

2.4 Scope, specific research objectives and limitations

Based on the descriptions given in Section 2.1 and Section 2.2, this section briefly specifies the research objectives, the intended scope of the proposed framework and the scope of the case study shown for the specific SMS design approach outlined in Section 2.3.

The focus of the proposed framework in Section 3 is on evaluation; that is, it aims to serve as a basis for systematically detecting strengths and weaknesses in the SMS design and operational use. It is a plausible hypothesis that a positively evaluated SMS will lead to better safety outcomes, and hence a higher level of pragmatic validity. The specific objective is to propose an evaluation framework tailored to the STAMP-based design approach for maritime SMS as described in Valdez Banda and Goerlandt (2018). The case study on the SMS developed for VTS Finland, elaborated in Valdez Banda and Goerlandt (2017, 2018), aims to test the framework, providing insights into the positive and negative aspects of the designed SMS. The case study results and the experiences gained from executing the framework are also used to formulate further research directions.

The proposed framework is introduced as an initial framework: as no earlier work has been dedicated to this in the maritime domain, it is likely that the framework itself may require modification and extension. While the proposed SMS evaluation framework may be more broadly applicable in the sense that the underlying ideas can be adapted to SMS based on other theories and applying different tools, this is not further explored in the current work. In the case study, the focus is exclusively on SMS design evaluation, because this particular design elaborated in Valdez Banda and Goerlandt (2017, 2018) has not yet been implemented in the organization. Finally, validation (as defined in Section 2.2) is out of the scope of the current paper.

3 A proposed evaluation framework for safety management systems

3.1 Generic framework: foundations

Considering the generic model of SMS presented in Fig. 1, it is evident that an evaluation framework aligns well with the purpose of the auditing process and tools. Hence, in general terms, it should guide an evaluator to consider how well the SMS implements applicable standards and how well the adopted methods and tools support the analysis, assessment and safety-related decision-making in the given organizational context. The evaluation also needs to consider the underlying theoretical basis on which the SMS is conditioned.

Whereas the generic model is useful to understand the relevant issues in SMS design and operations at a theoretical, standards and practical level, it does not specify which components are included in an SMS. However, it is considered important to structure and formulate an evaluation framework to be more in line with practical components of the SMS itself. This is primarily because this leads to more natural communication between the evaluator and the designers and users of the SMS. In SMS evaluation frameworks in other industries, such as Basso et al. (2004), Teo and Ling (2006) and Chang et al. (2015), this has also been the approach.

Rather than focusing directly on the SMS components, the presented evaluation framework focuses on different phases in the evaluation process, representing the different functions that the evaluation serves. This is because the exact components included in an SMS may vary between particular instantiations, but still serve the same functions. Referring to the main dimensions of SMS identified by Fernández-Muñiz et al. (2007) across industries, outlined in the introduction, it is evident that those are

also reflected in maritime SMS. This is the case for SMS for shipping companies based on the ISM Code (Batalden and Oltedal 2018) and/or the Tanker Management Self-Assessment (TMSA) (Valdez Banda et al. 2016). It is also the case for SMS developed for the vessel traffic services, a maritime service provider whose main objectives are to provide information and navigational assistance and to organize maritime traffic (Valdez Banda and Goerlandt 2017, 2018). However, the exact components applied in these SMS vary, making it a poor basis for structuring the evaluation framework.

The evaluation framework for SMS design and operation presented here is structured around four phases, presented in Fig. 2. In the context of the generic model by Li and Guldenmund (2018) presented in Fig. 1, phase A is mostly situated between the standards and practical levels. Phases B and C primarily address the practical level from the viewpoint of methods and techniques, but also link to the standards and theoretical level. Phase D is situated mostly at the practical level, focusing on internal auditing tools, which links to the methods and techniques for SMS performance evaluation.

Phase A focuses on the validation of the SMS support to the safety management policy. This phase includes six major clusters for policy analysis presented in Mayer et al. (2004). The first cluster is *research and analyse*, a perspective on policy analysis as knowledge creation. The second cluster is *design and recommend*, the translation of new knowledge into new policy design. The third cluster is *clarify arguments and values*, the normative and ethical questions and opinion behind policy. The fourth cluster is *provide strategic advice*, the development of the most effective strategy for achieving the policy goals. The fifth cluster is *democratize*, the development of equal access to the policy process for all stakeholders. The sixth cluster is *mediate*, the negotiation in policymaking and the interaction involved and progress in that process. Based on these clusters, Yücel and van Daalen (2009) structured specific actions for evaluating the policy support.

In phase A, these actions evaluate the analytical capability, advisory capability, strategic capability, mediation capability, participatory capability and argumentation capability of the SMS to support the policy. Table 1 introduces the general queries used to evaluate the system support to the safety management policy.

Phase B focuses on the evaluation of the usability of the system. This phase focuses on the analysis of the validity of the possible uses of the SMS as presented in Hodges (1991). The analysis focuses on the evaluation of six main clusters. The first cluster is

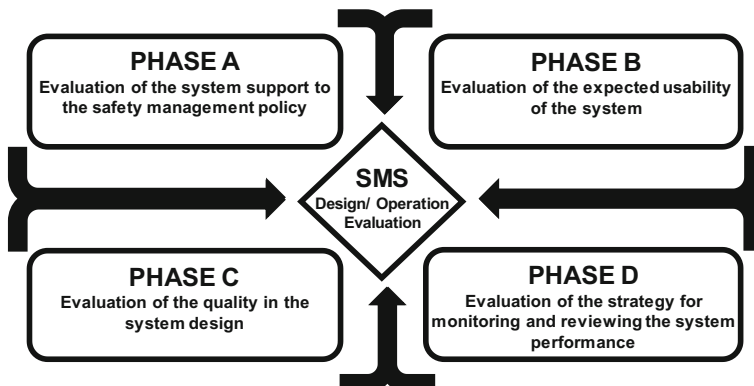


Fig. 2 The four phases of the SMS design/operation evaluation

Table 1 The queries and the aspects reviewed to evaluate the capabilities of the design and operation of the SMS

Code	Capability	Queries	SMS design	SMS operation
Phase A				
A.1	Analytical	Is the SMS a good reflection of the real system?	Evaluating whether the design of the SMS can accurately represent the safety management practices of the organization.	Evaluating whether the operation of the SMS accurately represents the safety management practices of the organization.
A.2	Advisory	Is the SMS a good guide for policy application?	Evaluating whether the designed SMS can serve as a guide to execute the safety policy of the organization and promote the organizational policy strategy.	Evaluating whether the SMS serves as a guide to execute the safety policy of the organization and as a guide to achieve the safety management strategy of the organization.
A.3	Strategic	Is the SMS supportive for leading the policy?		
A.4	Mediation	Is the SMS a good mediator among different stakeholders?	Evaluating whether the SMS design has the adequate structure to support and promote cooperation with the system stakeholders. Evaluating whether the SMS design provides means to analyse safety management issues with external stakeholders.	Evaluating whether the SMS operation contributes to the support and promotion of cooperation with the system stakeholders. Evaluating whether the SMS operation is suitable for analysing safety management issues with external stakeholders.
A.5	Participatory	Is the SMS a good promoter of personnel participation?	Evaluating whether the design of the SMS is suitable for promoting personnel involvement in the actual management of safety.	Evaluating whether the SMS operation supports personnel involvement in the actual management of safety.
A.6	Argumentation	Is the scope of the SMS clear?	Evaluating whether the SMS covers the majority of the crucial elements linked to the safety management policy of the organization. This aims at collecting feedback to update the design of the SMS.	Evaluating whether the SMS covers the majority of the crucial elements linked to the safety management policy of the organization. This aims at collecting feedback after the operation of the SMS.
		Is the SMS handling relevant information?	Evaluating whether the SMS design uses relevant information for the management of the policy.	Evaluating whether the SMS operation supports the management of essential safety information.
		Is the SMS linked to the policy practices?	Evaluating whether the SMS design is adequate	Evaluating whether the operation of the SMS

Table 1 (continued)

Code	Capability	Queries	SMS design	SMS operation
			to generate controls for accident prevention.	resulted as planned in the design phase. Evaluating whether the SMS prevents accidents.
		Is the SMS good for knowledge creation?	Evaluating whether the SMS design can generate and manage knowledge to link the theoretical foundations of the SMS and the safety management practices of the organization.	Evaluating whether the operation of the SMS contributes to the theoretical understanding of the safety management practices.
Phase B				
B.1	Value promotion	Does the SMS provide guidance and promote the values of the organization?	Evaluating whether the design of the SMS has properly adopted the values of the safety management strategy of the organization.	Evaluating whether the SMS supports the installation and development of the safety culture.
B.2	Behaviour induction	Does the SMS positively induce employees' behaviour?		
B.3	Efficacy	Does the SMS appropriately balance costs and benefits?		Evaluating whether the SMS is functional and cost-effective.
Phase C				
C.1	Functionality	Is the SMS suitable for the compliance of the demands from the organization and regulations?	Evaluating whether the SMS design can fulfil the internal demands in the organization and demands in applicable safety regulations.	Evaluating whether the SMS has the expected quality in the operation. This includes the results from the SMS evaluation (e.g. internal and external audits)
C.2	Reliability	Is the SMS reliable?	Evaluating whether the SMS design is ready for implementation.	Evaluating whether the operation of the SMS is reliable.
C.3	Usability	Is the SMS easy to learn?	Evaluating whether the SMS design is clearly understood at all levels in the organization.	Evaluating whether the SMS has the correct usability (the SMS is used as planned).
C.4	Maintainability	Is the SMS easy to maintain and update?	Evaluating whether the SMS is suitable for creating a maintenance plan that ensures its functionality.	Evaluating whether the SMS is easy to maintain.
Phase D				
D.1	Alignment	Is the SMS strategy for monitoring and reviewing the SMS	Evaluating whether the mechanisms designed to monitor and review	Evaluating whether the mechanisms used for monitoring and

Table 1 (continued)

Code	Capability	Queries	SMS design	SMS operation
		performance aligned with organizational practice?	the system performance match the ones used in the organization.	reviewing the system performance match the practices adopted in the organization.
D.2	Adoptability	Are the variables utilized to review and monitor the SMS performance similar to those utilized in practice?	Evaluating whether the content needed for the functioning of the adopted mechanism covers the elements of the safety management in the organization.	

the use of the SMS as a *bookkeeping device*, the provision of means to improve data quality and processing. The second cluster is the use of the SMS as *part of an automatic management system*, the provision of outputs in an automatic function. The third cluster is the SMS as *vehicle for a fortiori arguments*, the creation of a flexible and responsive system. The fourth cluster is the use of the SMS as an *aid to thinking and hypothesizing*, the provision of assumptions that produce knowledge. The fifth cluster is the use of the SMS as an *aid in selling an idea*, the conveying of aspects of the system into concrete actions. The sixth cluster is the use of the SMS as a *training aid to induce behaviour*, the induction of desired behaviour.

In phase B, specific actions are considered to evaluate the capability of the SMS to promote value in the organization, to induce behaviour and to determine efficacy levels. Table 1 introduces the general queries used to evaluate the usability of the system.

Phase C focuses on the evaluation of quality in the SMS design. The structure of this phase is based on the analysis of the quality perception by SMS users described in Bevan (1999). This focuses on the evaluation of six main clusters. The first cluster is the *functionality* of the SMS; this cluster covers the evaluation of the accuracy, suitability, interoperability and compliance of the SMS. The second cluster is the *reliability* of the SMS; this cluster covers the maturity, fault tolerance and recoverability of the SMS. The third cluster is the *usability* of the SMS; this cluster covers the understandability, learnability and operability of the SMS. The fourth cluster is the *efficiency* of the SMS; this cluster covers time, resource and utilization of the SMS. The fifth cluster is the *maintainability* of the SMS; this cluster covers the analysability, changeability, stability and testability of the SMS. The sixth cluster is the portability of the SMS; this cluster covers the adaptability, conformance and replaceability of the SMS.

In phase C, specific actions are considered to evaluate the capability of the SMS functionality, reliability, usability and maintainability. Table 1 introduces the general queries used to evaluate quality in the SMS design.

Phase D focuses on the evaluation of the strategy for monitoring and reviewing SMS performance, particularly the means utilized to monitor and review the SMS. This phase is adapted from the framework for measuring the latent and dynamic variables of an SMS presented in Pitchforth and Mengersen (2013). The framework includes seven main clusters for system validity: nomological validity, face validity, content validity, concurrent validity, predictive validity, convergent validity and discriminant validity. In this phase D, face validity and content validity are two clusters considered for

evaluating the strategy for monitoring and reviewing SMS performance. Face validity focuses on evaluating the structure of the SMS to adopt the strategy for monitoring and reviewing its performance. Content validity focuses on identifying relevant factors for monitoring and reviewing SMS performance.

In phase D, specific actions are considered to evaluate the capability of the SMS strategy to be aligned with organizational practices, and the capability of the SMS strategy to adopt the instruments currently used for monitoring and reviewing safety management performance. Table 1 introduces the general queries used to evaluate the strategy for monitoring and reviewing SMS performance.

3.2 Framework structure: SMS design

3.2.1 Evaluation of the system support to the safety management policy

Phase A of the framework focuses on evaluating the design of the SMS in terms of its support for implementing the safety policy of the organization and the estimation of the SMS contribution to achieving the safety management objectives of the organization. In this evaluation of the SMS design, the six listed capabilities of the SMS are evaluated. Table 1 presents the aspects evaluated in the SMS design for phase A.

3.2.2 Evaluation of the expected usability of the system (based on design features)

Phase B of the framework focuses on evaluating the design of the SMS in terms of the expected usability of the system. It represents an anticipated evaluation of the support that the SMS can provide to install the intended safety culture in the organization. In this evaluation of the SMS design, only one capability can be evaluated at this phase (value promotion). Table 1 presents the aspects evaluated in the SMS design for phase B.

3.2.3 Evaluation of quality in the system design

Phase C of the framework focuses on evaluating the perceived quality of the system design and the expected quality of the system functionality. It evaluates the design of the system in terms of expected quality during its life cycle, its expected reliability, usability and maintainability. Table 1 presents the aspects evaluated in the SMS design for phase C.

3.2.4 Evaluation of the strategy for monitoring and reviewing the system performance

Phase D of the framework focuses on evaluating the adopted strategy for monitoring and assessing the system performance. This refers to the evaluation of the mechanisms (e.g. process, tools and applications) implemented to monitor and review the system performance and to determine actions for guiding the system performance into defined safety margins. Table 1 presents the aspects evaluated in the SMS design for phase D.

3.3 Framework structure: SMS operation

The evaluation of the SMS design is the initial step in the framework. The second step is the evaluation of the actual operation of the system. This aims at representing the

actual effectiveness of the SMS in reaching its general objective. Using the four phases of the design evaluation, particular aspects are listed to evaluate the capabilities of the SMS operation. Table 1 presents the particular aspects of SMS operation evaluated in the four phases.

4 Case study: VTS Finland SMS

VTS Finland is one of the main actors responsible for monitoring and controlling the safety and smooth development of maritime traffic in Finnish sea areas (Praetorius et al. 2015). The proposed framework in this study is applied to evaluate a new SMS designed for VTS Finland introduced in Valdez Banda and Goerlandt (2018). This new SMS aims at representing the safety function of VTS Finland and the controls utilized to ensure their internal safety management and the safety of the navigation in Finnish sea areas. The SMS contains 13 general requirements that rule the three main objectives of VTS Finland. The SMS focuses on the prevention of 20 accident scenarios and the mitigation of 26 identified hazards in the functioning of VTS. The SMS includes a tool for monitoring and reviewing the system performance. The tool runs with the application of 31 key performance indicators (KPIs) distributed among the 13 system requirements. These KPIs guide and reinforce the requirements of the system, monitor system functioning and present information about temporary results of system activity. The details of the accident scenarios, hazards, the SMS general requirements and their KPIs are presented in Tables 1, 2 and 4 and Appendix E in Valdez Banda and Goerlandt (2018).

4.1 Evaluation framework application

4.1.1 Expert workshop

The application of the framework was carried out in a workshop with experts in the provision of VTS. This group consists of ten experts, including one manager, two supervisors and seven officers of VTS Finland. The managers have more than 5 years of experience in the function of VTS in Finland. The supervisors have almost 10 years of experience in the function of VTS and more than 10 years of experience onboard vessels (ship bridge operations). The officers have from 2 to 6 years of experience in the function of VTS. All the officers have practical experience onboard ships.

The workshop was organized in four sessions. The first session presented the foundations and the general structure of the process for designing the SMS for VTS Finland. The second session presented the purpose, scope and objectives of the SMS. The third session presented the actual SMS, including the accidents and hazards that the SMS aims to prevent and mitigate, the requirements and the assumptions of the SMS and the logic principle of the requirements of the SMS. The fourth session presented the description of the KPIs and the tool designed to monitor and review the performance of the SMS. This includes the description of the method used to define the KPIs, the description of the 31 KPIs and the functionality of the performance monitoring and reviewing tool.

The presentations at each stage have two purposes:

- Present information to evaluate the SMS with the application of a questionnaire for each phase of the framework. The details of this questionnaire are presented in Appendix Table 2.
- Collect (via note recording) feedback about any aspect included in the SMS design.

4.1.2 Questionnaire to assess the phases of the evaluation framework

Each aspect included in the four phases of the evaluation framework has a set of questions for the experts participating in the workshop. The majority of these questions are answered with one of the following options: completely disagree, disagree, neither agree nor disagree, agree, completely agree. A few open questions are also included in the questionnaire. The questions assess the capabilities (see Table 1) of the design and operation of the SMS. Appendix Table 2 presents the questions utilized to evaluate the design of the VTS Finland SMS. The evaluation only concerns the design phase of the SMS, not the operational phase of the SMS.

5 Results

5.1 Results of the questionnaire to evaluate the capabilities of the SMS designed for VTS Finland

Figure 3 presents the answers to the questions in phase A collected from the ten experts participating in the workshop. Figure 4 presents the answers to the questions in phases B, C and D collected from the ten experts participating in the workshop.

5.2 General feedback on the SMS design

The questionnaire included two open questions (A5.3 and A6.2). The first asks the respondent to mention aspects that should be considered in promoting personnel involvement. The experts pointed out the need to define the exact role of VTS personnel in the implementation and development of the SMS. They also remarked that the SMS design must ensure that the SMS information reaches all levels of the organization. The experts highlighted the importance of carrying out an initial launch of the SMS and a training campaign for the VTS personnel. Finally, the experts mentioned that the SMS should be efficiently linked to the equipment of VTS centres. The second question is about the aspects of the SMS that have to be either included or removed from this initial design. The experts did not mention any such aspects in this open question.

Feedback was collected by recording notes and it provided different aspects that need to be considered in the design of the SMS. The first aspect is the frequency with which VTS officers need to interact with vessels in traffic. The experts mentioned that an officer establishes interaction with vessels in traffic about 6000 times per year. This interaction is mainly about information sharing, focused primarily on proactive safety work such as providing information about outbound traffic for vessels entering an area, information

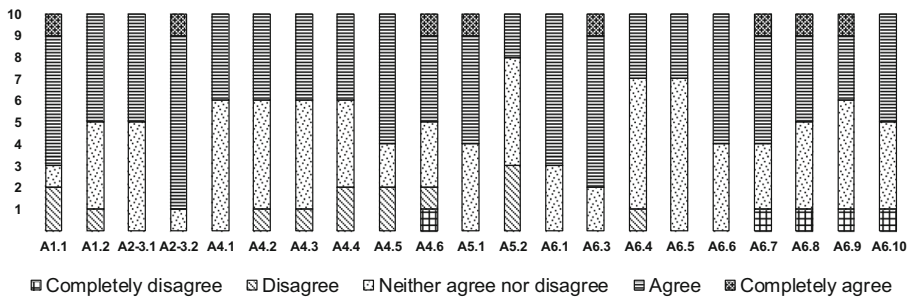


Fig. 3 Answers provided to the questionnaire of evaluation phase A in the workshop

about possible congestions or dangers ahead, information about piloting (e.g. delayed pilot arrival) and information about ice status. The experts mentioned that they also maintain active communication to prevent near misses and to ensure that ships follow the rules. Particular topics were further discussed, including the coordination of piloting services and winter navigation. The experts mentioned that nowadays, there are more captains with a pilot licence in the VTS areas, and that the main role of VTS during winter navigation is to support icebreakers, maintain awareness and monitor the traffic.

The experts mentioned that most ship violations concern technical failures. These are reported by the officers and supervisors, and submitted to their reporting system. However, feedback about those reports is only received from some flag states. The experts discussed the need to report more proactive actions. This practice is currently adopted in their quality management system. Moreover, VTS Finland intensified the period of reporting to twice a year in order to understand VTS work as done and to minimize the burden of reporting. Finally, the experts mentioned that VTS Finland carries out periodic performance evaluations with external stakeholders. However, they would prefer to receive this feedback more constantly and in a more systematic form.

6 Discussion

6.1 Results of the case study for design evaluation of SMS for VTS Finland

Referring to Section 3.1, *phase A* focuses on the evaluation of the SMS design support for the safety management policy. Experts found that the system design and structure accurately represent the actual safety management at VTS (A1). However, only 50% of

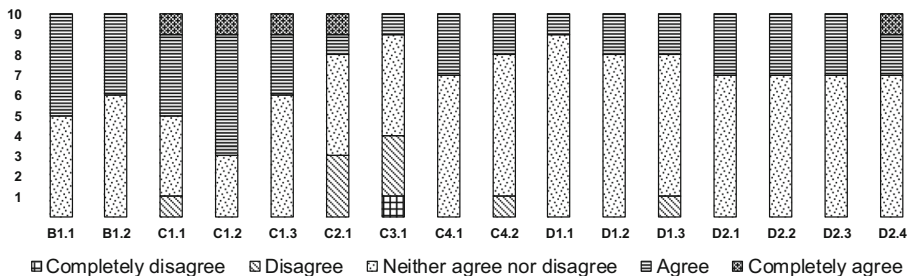


Fig. 4 Answers provided to the questionnaire of evaluation phases B, C and D in the workshop

the experts think that it is ready to be applied immediately into practice, while 40% of the experts are neither in favour nor against it, and only 10% think it is not ready. The survey feedback indicates that the foundations of the proposed SMS are difficult to understand in one session. The experts mentioned that specialized training and practical exercises could enhance the effectiveness of single-session training.

The experts consider that the SMS has potential to improve the management of safety at VTS (A2); 50% of the experts agree that the SMS represents a means to achieve the safety management strategy. In the evaluation of the SMS capability to support and promote cooperation with the SMS stakeholders (A3), 40% see this clearly represented in the designed structure. However, the other experts indicated that this aspect cannot be considered properly until the SMS is used in actual operation.

The experts mostly agree that the SMS is a good representation of the function of VTS and that it serves as a good basis to discuss safety management issues with external stakeholders (A4). The feedback collected indicates strong support for including means to support cooperation with external stakeholders in the SMS. The experts give a positive evaluation to the SMS designed for this task. However, the experts also indicated that the SMS should more clearly represent how the external stakeholders are intended to be involved in SMS implementation during operation.

The evaluation of the potential of the designed SMS to promote personnel involvement in organizational safety management (A5) identified different issues. The experts agree that the designed SMS can promote and improve personnel involvement. However, the experts pointed out that the structure does not include a clear representation of personnel involvement. The collected feedback clearly expresses that the SMS should reach all levels in the organization and align the reporting requirements for the SMS with the operational tasks for performing the VTS functions.

The expert evaluation indicates that the SMS scope includes most of the crucial elements of VTS Finland's safety management policy (A6). No further feedback was given regarding missing or unnecessary aspects in the SMS design.

The experts evaluated that the SMS includes relevant information to ensure the safety of VTS and safety in the maritime operational context for which VTS provides services (A7). However, only 30% of the experts agree the SMS represents a good means to share information with VTS stakeholders. This highlights again the need for a clearer representation of the involvement of external stakeholders in the functioning of the system.

The experts evaluated that the SMS features a good design to detect and anticipate potential hazards affecting the functioning of VTS (A8). In evaluating the capability of the SMS to support the generation of knowledge and skills that strengthen the link between the SMS foundations and the actual safety management practices in the organization (A9), 60% of the experts agree that the SMS can strengthen the knowledge and skills of VTS personnel. Although 50% agree the SMS provides support to the general strategy, 10% entirely disagree with this claim. A deeper analysis is required in order to identify the issues behind this disagreement.

Referring to Section 3.1, *phase B* addresses the expected usability of the SMS in the proposed evaluation framework. Here, 50% of the experts agree the SMS is properly adapted to the safety management strategy of VTS Finland (B1). The other 50% are neither in favour nor against it. Similarly, 40% of the experts expect that, with the current design, the system can support the development of the safety management strategy. While 40% of the experts agree with the expectation that the SMS would function cost-

effectively in the organization, 10% of the experts entirely disagree with this expectation. These aspects are important to consider in preparing the SMS design for the implementation phase. Targeted clarifications are needed to describe how the SMS supports the development of a safety management strategy and to convince people about this.

Phase C, according to Section 3.1, focuses on evaluating the quality of the SMS. It initially evaluates how well the SMS incorporates organizational and regulatory requirements (C1). The majority of the experts agree that the SMS is accurate and suitable for the management of safety at VTS. In addition, 40% of the experts agree the SMS properly covers the regulatory demands. The rest of the experts mentioned that they need to better study the design foundations to assess this aspect.

Based on the evaluation of the quality of the SMS design, only 20% of the experts agree that the SMS is ready for implementation (C2). This is one of the aspects with a higher disagreement percentage (30%), which clearly indicates that the experts disagree that the SMS should be implemented in its current version. The other aspect with higher disagreement (40%) is the ease of understanding the system (C3), and feedback indicates that clarifications are required for understanding the overall SMS design, the meaning and use of the KPIs, and how those are linked with everyday operational work. The experts proposed that an initial launching of the SMS should be accompanied with a training campaign in order to enable understanding the system design and how it should be used in an operational setting. Only 30% of the experts agree the SMS is easy to maintain, while only 20% agree that the SMS and its content are easy to update (C4). This negative view is likely associated with the perceived difficulty in understanding the SMS design.

Referring to Section 3.1, *phase D* addresses the evaluation of the strategy for monitoring and reviewing the system performance. This initially evaluates the mechanisms and tools for monitoring and reviewing the system performance (D1). Expert feedback pointed out that the use of a Bayesian Network-based tool is an interesting approach with a complex functionality. However, only 10% of the experts agree that the tool is adequate for monitoring and reviewing the system. For the other experts, this is unclear. They mentioned that perhaps presenting a simplified version of the tool, showing how KPIs that are already familiar to them are processed, can give a better understanding of how the tool functions.

Finally, the experts evaluated whether the content represented in the Bayesian Network tool, i.e. the proposed KPIs, covers relevant aspects for assessing the system safety and for guiding the decisions. In this task, 30% of the experts agree that the proposed KPIs are relevant for these purposes. The other experts are neither in favour nor against it. However, they had a more positive perspective towards the tool process to plan KPIs, collect the information and report KPIs, and the overall functionality of the tool.

On balance, the application of the evaluation framework presents a generally rather positive reception of the SMS by the experts. In 13 instances, the experts completely agree with the claims on the foundations of the SMS design, related to aspects A1, A2, A4, A5, A7, A9, C1, C2 and D2. In 149 instances, the experts agree with the claims on the foundations of the SMS design with respect to all aspects considered in the evaluation framework. In 188 instances, the experts neither agree nor disagree with the claims on the foundations of the SMS design with respect to all aspects considered in the evaluation framework. In 23 instances, the experts disagree with the claims on the foundations of the SMS design related to aspects A1, A3, A4, A5, A7, C1, C2, C3,

C4 and D1. Only in seven instances, the experts entirely disagree with the claims on the foundations of the SMS design related to aspects A4, A9, B1 and C3.

Hence, the case study illustrates that the SMS design elaborated in Valdez Banda and Goerlandt (2017, 2018) is a good initial representation of an SMS for VTS Finland. Compared with existing maritime SMS design methods, e.g. Valdez Banda et al. (2016), the underlying design approach has merits in that it is based on a theoretical understanding on organizational accident occurrence. Qualitative feedback from the experts also indicates that they appreciate the fact that the KPIs are rooted in such a theoretical accident model as well as a hazard identification and analysis method which accounts for the specific functions represented in the VTS system. The developed KPIs were seen as plausible elements of the safety management function of the VTS, and the Bayesian Network tool was seen as an interesting approach to combining these KPIs to assess safety performance and guide decision-making.

However, the experts also clearly indicated they do not support the immediate implementation of the SMS into the VTS processes. The evaluation framework helps to raise issues needing improvement in the SMS design, including (i) the intended involvement of external stakeholders in the SMS during operation, (ii) the personnel involvement in the SMS, (iii) the relation between the SMS and the organizational safety management strategy and (iv) the representation of the KPIs in the BN tool.

In the qualitative feedback, the experts also indicated that they needed more time and explanations on how exactly the accident generating mechanisms in STAMP, and the hazard analysis method STPA, work. Similarly, while the evaluation generally indicates that relevant KPIs are identified and that the BN tool may be a good tool to combine these, the experts indicated that more time and practical exercises would be necessary to better understand what exactly the KPIs represent and how these link with the suggested decisions in the BN tool. This also relates to the finding that there is a strong need for training on the SMS contents and practical use if a decision were made to implement the designed SMS in actual organizational practice.

Considering the generally positive reception of the SMS design, even with the experts' lack of familiarity with STAMP, STPA and BNs, it is plausible to assume that with more time to understand the theoretical and methodological basis of the SMS design, more experts would be supportive of the implementation of (a somewhat modified) SMS design for VTS Finland.

This case study shows the utility of the proposed evaluation framework for assessing the overall value of the designed SMS and for guiding the further adaptations, clarifications and improvements before proceeding to the implementation phase.

6.2 Case study: limitations and future work

While the case study provides useful insights, it is important to reflect on the limitations of the case study and to provide guidance for future work.

One important issue concerns the setup of the expert workshop. For the purposes of the evaluation, the authors considered it important to gain feedback from a range of experts having different roles in the operation of VTS Finland. This is because it is known that SMS needs to be supported by all levels in the organization, from the management to the operator levels (Fernández-Muñoz et al. 2007; Kelly 2017). Therefore, a VTS manager, supervisors and operators were consulted, as outlined in

Section 4.1.1. Although VTS Finland provided excellent support to this research, it is an organization with high operational demands, and thus, there were limits to the amount of time that the expert participants could make available for this research. Furthermore, VTS Finland wanted to have one workshop in which all the interested experts could participate, rather than multiple events. This constrained the research format, first and foremost, in terms of the time available to explain the foundations of the SMS design method developed by Valdez Banda and Goerlandt (2018) and the proposed SMS design for VTS Finland. As found in the results of Section 6.1, this limited time is a contributing factor to the support for implementing the SMS in the organization. The aim is to create a simple and fully understood SMS that supports personnel's role and responsibilities and safety compliance (Antonsen et al. 2008).

The workshop setup also had important repercussions for how the authors could operationalize the evaluation framework as a data collection tool. Given the limited time available to collect feedback, it was considered best to use a survey format, as outlined in Section 4.1.2. This allows a broad and rather rough appreciation of the aspects of the SMS design needing most improvement, but it has limitations in understanding the reasons behind the answers. A group discussion was therefore also included in the workshop to highlight some of the reasons behind the scores. However, it is acknowledged that more focused individual and/or focus group interviews would provide much deeper insights into the experts' reasoning. However, time limitations did not allow such an approach.

Another issue in the case study is that it only considers the initial design of the SMS. From the interactions with the experts, especially the VTS operators, it was clear that there are important concerns related to the use of the SMS in operational practice. Some SMS aspects relevant for operations could be identified in the design evaluation stage, e.g. the intended involvement of personnel and external stakeholders (see Section 6.1). It is clearly important to carefully further consider these aspects in the SMS design stage, but even more to implement appropriate processes to establish and maintain the personnel support for the SMS during operation. As the SMS has not been implemented yet, it is evidently not possible to test how well the SMS actually aligns with the personnel expectations and needs.

For future work, the priorities are therefore to execute selected in-depth interviews with selected experts to better understand the reasons for the evaluations, especially the more negative ones. There is also a need to develop better educational materials focusing on the underlying theoretical aspects of the SMS design, in particular the STAMP theory, the STPA method and the BN tool, which is expected to lead to further expert support for the SMS design. This closely links to the need to develop training materials for personnel at all organizational levels to clarify roles and procedures in the practical use of the SMS. Finally, more long-term work is needed to elaborate the evaluation framework to cover more operational issues related to the practical use of SMS.

6.3 Framework: limitations and future work

The evaluation framework proposed in this work is explicitly presented as an initial evaluation framework. The four phases outlined in Section 3.1 are associated with different levels and elements of the generic SMS model by Li and Guldenmund (2018), providing some justification that the framework covers relevant aspects. The contents of

the phases are based on evaluation processes addressing policy (phase A), usability (phase B), quality (phase C) and monitoring and reviewing system performance (phase D). These phases cover similar issues as in evaluation frameworks presented for other industries, e.g. Basso et al. (2004), Teo et al. (2006) and Chang et al. (2015). The evaluation framework presented here, however, differs in the approach adopted, as we do not focus on the SMS components directly, but rather on more generic aspects of an SMS that are known to be relevant based on the generic model by Li and Guldenmund (2018).

The approach underlying the presented evaluation framework focuses more directly on the theoretical basis and the methodological implementation of the SMS, which is the focus of the SMS design approach by Valdez Banda and Goerlandt (2018). This more fundamental underlying theoretical basis sets this design approach apart from existing work on maritime SMS, as outlined in the introduction. This distinction with other work on maritime SMS also justifies why a specific evaluation framework is needed for the new SMS design approach. As it does not directly focus on SMS components (e.g. required by regulations or standards), but rather on the underlying accident causation mechanisms, system hazards and safety controls, another method is needed to evaluate whether an instantiation of a specific SMS design is considered appropriate for practical implementation. It is explicitly not the intention to evaluate the designed SMS exclusively based on compliance with the requirements imposed by standards, which however is the standard practice in many maritime SMS (Batalden and Oltedal 2018, Lappalainen 2016). Therefore, a focus on phases covering the SMS functions described by Li and Guldenmund (2018) offers an alternative approach that focuses more on the mechanisms underlying the SMS.

The framework is intended primarily to cover the aspects considered in the design approach by Valdez Banda and Goerlandt (2018). The focus in the evaluation framework on the contents and structure of the SMS, through the KPIs and their relation to decisions, is rooted in the importance of the underlying accident theory (STAMP); its associated hazard identification and analysis method (STPA) and the organizational management theory known as the Deming cycle (plan-do-check-act). The focus on usability and quality also relates to the SMS structure (as resulting from the STPA analysis), but furthermore considers the tools and techniques for practically implementing the SMS, as indicated by Li and Guldenmund (2018). Finally, the focus on evaluating how well the SMS covers the safety management policies relates to the standards level (*ibid*).

The four phases of Section 3.1 hence cover relevant aspects of the standards, practical and theoretical level indicated in Fig. 1, in line with the theories, techniques and tools proposed in the SMS design approach by Valdez Banda and Goerlandt (2018). However, the authors are aware that additional theories and methods are likely to be necessary to more fully evaluate the complete SMS performance. These additional evaluations may relate, for instance, to the techniques for designing checklists used as a basis for numerically measuring the KPIs in the SMS, see e.g. (Rae 2018). Another example may relate to underlying theories of organizational behaviour, e.g. (Allaire and Firsirotu 1984), which may be used to evaluate how the SMS is actually used in operational practice, as opposed to the intended use in the design stage.

Hence, the proposed evaluation framework likely is incomplete in the sense that other techniques and theories will underlie its actual operational use. The framework itself would benefit from further academic scrutiny, and possibly revision and updating. The

proposal is therefore an initial framework, covering the aspects included in Valdez Banda and Goerlandt (2018) as the fundamental SMS design base. It is left for future research to extend the evaluation framework with additional aspects covering additional techniques and theories underlying its actual implementation in operational practice. At the current stage of development of a theoretically rooted maritime SMS, the authors however feel confident that this initial framework is a useful step. We hope that this initial framework can guide the research field towards a focus on theories of system safety and techniques to operationalize and implement these in practice. Given that the current state of the art in maritime SMS focuses almost exclusively on the SMS components as mandated by regulations or standards (Batalden and Oltedal 2018; Lappalainen 2016), whereas there are strong arguments for more explicitly considering system safety theories (Li and Guldenmund 2018; Leveson 2011), the authors consider the initial framework proposed here a meaningful contribution to shift the focus from compliance with regulations to active engagement with system safety theories and methods.

7 Conclusions

This study presents an initial evaluation framework for design and operational use of safety management systems. The framework is guided by a process to evaluate the SMS design and operation in four phases, rooted in the theoretical basis and methodological implementation in line with the different levels considered in a generic model for SMS. Phase A evaluates the SMS support to the safety management policy of the organization. Phase B evaluates the usability of the SMS. Phase C evaluates the quality of the system. Phase D evaluates the strategy adopted for monitoring and reviewing system performance. In this study, these four phases of the framework are focused on evaluating the SMS design prior to the implementation phase.

The framework is applied to evaluate an SMS designed for VTS Finland. The aim is to represent the status of the system design and to detect potential weaknesses that need to be corrected, clarified or updated before its implementation. In the case study, the application of the evaluation framework showed that the designed SMS received an overall positive evaluation in terms of how accurately it represents safety management at VTS. Similarly, VTS personnel evaluated that the SMS has potential to improve the safety management of the organization. Other positive aspects include the complete and relevant information covered by the SMS and the expected suitability of the system for the management of safety.

The evaluation suggests that the main weaknesses in the designed SMS concern the description of the personnel involvement within the system and the complexity of the system. The experts stressed that before executing any action for implementing the system, they would need to more clearly understand the underlying theoretical basis of the SMS design, and especially the functioning of the SMS in relation to their actual work practices. The experts indicated a need for preparing an SMS training programme for all the VTS personnel. The evaluation indicated that the SMS is not yet matured enough to be implemented.

In general, the proposed framework provides a systematic approach to evaluate key aspects of the design, implementation and operation of SMS. The application of the framework to the SMS designed for VTS Finland provides an informative evaluation

process that represents the status of the system design in terms of readiness to proceed with its implementation. Nevertheless, it is acknowledged that the proposed evaluation likely needs further extension, especially related to theories of organizational behaviour that may point to specific issues for the operational use of the SMS and to evaluation methods for additional techniques, methods and tools for practically using the SMS. Correspondingly, several directions for future research have been indicated.

Acknowledgements Open access funding provided by Aalto University. The work presented in this study started in the research project “Strategic and Operational Risk Management for Wintertime Maritime Transportation System” (BONUS STORMWINDS) and finalized in the project “Reliability and Safety Engineering and Technology for large maritime systems” (RESET). BONUS STORMWINDS received funding from BONUS, the joint Baltic Sea research and development programme (Art 185), funded jointly from the European Union’s Seventh Programme for research, technological development and demonstration and from the Academy of Finland. RESET is partially supported by the European Union’s Horizon 2020 Research and Innovation Programme RISE under grant agreement no. 730888 (RESET). The contributions by the second author are in part supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Canada Research Chairs programme.

Appendix

Table 2 The questionnaire to evaluate the capabilities of the SMS designed for VTS Finland

Code	Capability	Question
Phase A		
A.1	Analytical	A1.1. The SMS (with the current design structure) is a good representation of safety management in VTS. A1.2. The SMS has an adequate structure for applying it in practice.
A.2	Advisory	A2-3.1. The SMS is a good tool for achieving the VTS safety management strategy.
A.3	Strategic	A2-3.2. The SMS has potential to improve the management of safety in VTS.
A.4	Mediation	A4.1. The SMS represents an instrument to support internal cooperation and communication in VTS. A4.2. The structure of the system promotes and supports cooperation and communication with external stakeholders A4.3. The SMS promotes the harmonization of safety management practices internally and externally. A4.4. The system is adaptable to reach joint targets with all stakeholders, including shipping companies, pilots, SAR services, icebreakers and maritime authorities. A4.5. External stakeholders involved in the function of VTS are correctly represented in the structure of the SMS. A4.6. The SMS is a good mean to analyze issues regarding the function of VTS and the connection with external stakeholders.
A.5	Participatory	A5.1. The SMS can promote and/or improve the involvement of all VTS personnel in the safety management of the organization. A5.2. The structure of the SMS is adequate to ensure personnel involvement at all levels A5.3. Mention the aspects that should be considered for promoting personnel involvement. These are aspects to be included in the design of the SMS (open question).
A.6	Argumentation	A6.1. The scope of the SMS is adequately describing the functioning of VTS Finland. A6.2. Mention aspects of the SMS that have to be either included or removed (open question).

Table 2 (continued)

Code	Capability	Question
		A6.3. The SMS covers the main information for ensuring the safety of VTS centers and ship traffic.
		A6.4. The SMS represents a mean to transfer information among all system stakeholders.
		A6.5. The SMS provides empirical information generated by its functioning.
		A6.6. The SMS serve as tool to anticipate to potential hazards for the functioning of VTS.
		A6.7. The SMS can strengthen the knowledge and skills of VTS personnel.
		A6.8. The SMS can clearly transmit the importance of the safety management for VTS.
		A6.9. The SMS is functional to detect the most relevant aspects for safety management.
		A6.10. The SMS is supportive to plan the organizational strategy
Phase B		
B.1	Value promotion	B1.1. The SMS is adapted to the strategy at VTS. B1.2. The SMS is functional and supportive for the strategy.
Phase C		
C.1	Functionality	C1.1. The SMS is accurately representing the management of safety at VTS. C1.2. The SMS is suitable for the management of safety at VTS. C1.3. The SMS seems to cover the demands on regulations.
C.2	Reliability	C2.1. The current structure of the SMS is mature enough to be applied.
C.3	Usability	C3.1. The SMS and its content are easy to understand.
C.4	Maintainability	C4.1. The SMS is suitable for making the planning of its maintenance. C4.2. The SMS content is easy to update.
Phase D		
D.1	Alignment	D1.1. The tool structure (variables, variable labels and connections between them) is adequate for monitoring and reviewing the VTS safety performance. D1.2. Each variable of the tool includes appropriate states for the description of the system performance. D1.3. The parameters of the variables are similar to what the experts would expect.
D.2	Adoptability	D2.1. The structure of the tool contains all and only factors and relationships relevant to the tool output. Thus, the KPIs and actions cover relevant factors to ensure and improve safety. D2.2. Each variable of the tool contains only relevant states. D2.3. The discrete states of the variables are dimensionally consistent. The states have a good scale to represent the status of the requirements and the entire SMS. D2.4. The parameters of the input variables are easy to plan, including the collection of information to assess the system performance.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akyuz E, Celik M (2014) A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. *Saf Sci* 68:169–179
- Allaire Y, Firsirotu ME (1984) Theories of organizational culture. *Organ Stud* 5(3):193–226
- Antonsen S, Almklov P, Fenstad J (2008) Reducing the gap between procedures and practice lessons from a successful safety intervention. *Saf Sci Monit* 12(1):1–16
- Aven T (2014) What is safety science? *Saf Sci The Foundations of Safety Science* 67:15–20. <https://doi.org/10.1016/j.ssci.2013.07.026>
- Basso B, Carpegna C, Dibitonto C, Gaido G, Robotto A, Zonato C (2004) Reviewing the safety management system by incident investigation and performance indicators. *J Loss Prev Process Ind* 17(3):225–231
- Batalden B-M, Olstedal HA (2018) Safety management systems. In: Olstedal HA, Lützhöft M (eds) *Managing maritime safety*. Routledge, Abingdon
- Batalden B-M, Sydnes AK (2014) Maritime safety and the ISM Code: a study of investigated casualties and incidents. *WMU J Marit Aff* 1(13):3–25
- Batalden B-M, Sydnes AK (2015) Auditing in the maritime industry: a case study of the offshore supply vessel segment. *Saf Sci Monit* 19(1), article 3
- Bevan N (1999) Quality in use: meeting user needs for quality. *J Syst Softw* 49:89–96. [https://doi.org/10.1016/S0164-1212\(99\)00070-9](https://doi.org/10.1016/S0164-1212(99)00070-9)
- Bhattacharya S (2012) The effectiveness of the ISM Code: a qualitative enquiry. *Mar Policy* 36(2):528–535
- Bieder C, Bourrier M (2013) Tapping safety into rules: how desirable or avoidable is proceduralization? Ashgate Publishing Co., Farnham
- Celik M (2009) Designing of integrated quality and safety management system (IQSMS) for shipping operations. *Saf Sci* 47:569–577
- Chang JI, Liang C-L (2009) Performance evaluation of process safety management systems of paint manufacturing facilities. *J Loss Prev Process Ind* 22:398–402. <https://doi.org/10.1016/j.jlp.2009.02.004>
- Chang Y-H, Shao P-C, Chen HJ (2015) Performance evaluation of airport safety management systems in Taiwan. *Saf Sci* 75:72–86
- Dekker SWA (2014) The bureaucratization of safety. *Saf Sci* 70:348–357. <https://doi.org/10.1016/j.ssci.2014.07.015>
- Fernández-Muñiz B, Montes-Peón JM, Vázquez-Ordás CJ (2007) Safety management system: development and validation of a multidimensional scale. *J Loss Prev Process Ind* 20(1):52–68
- Gerede E (2015) A study of challenges to the success of the safety management system in aircraft maintenance organizations in Turkey. *Saf Sci* 73:106–116. <https://doi.org/10.1016/j.ssci.2014.11.013>
- Goerlandt F, Khakzad N, Reniers G (2017) Special issue: risk analysis validation and trust in risk management. *Saf Sci* 99:123–126. <https://doi.org/10.1016/j.ssci.2017.07.012>
- Hale A (2005) Safety management, what do we know, what do we believe we know, and what do we overlook. *J Appl Health Sci* 2005
- Hale A (2014) Foundations of safety science: a postscript. *Saf Sci* 67:64–69. <https://doi.org/10.1016/j.ssci.2014.03.001>
- Hale AR, Heming BHH, Carthey J, Kirwan B (1997) Modelling of safety management systems. *Saf Sci* 26(1–2):121–140
- Hodges JS (1991) Six (Or So) Things you can do with a bad model. *Oper Res* 39:355–365. <https://doi.org/10.1287/opre.39.3.355>
- Hollnagel E, Nemeth CP, Dekker S (2008) *Resilience engineering perspectives: remaining sensitive to the possibility of failure*. Ashgate Publishing, Ltd, Farnham
- ICAO (2009) *Safety management manual*, 2nd edn. International Civil Aviation Organization, Montreal
- IMO (1993) *International Management Code for the safe operation of ships and for pollution prevention*, International Safety Management (ISM) Code. IMO Document Res. A.741(18), Adopted November 4, 1993. London, United Kingdom
- Karakasnakli M, Vlachopoulos P, Pantouvakis A, Bouranta N (2018). ISM Code implementation: an investigation of safety issues in the shipping industry. *WMU J Marit Aff* 17(3):461–474
- Kelly T (2017) The role of the regulator in SMS. International Transport Forum. Discussion Paper 2017-7. International Transport Forum and Organization for Economic Co-operation and Development
- Kongsvik TØ, Størkersen KV, Antonsen S (2014) The relationship between regulation, safety management systems and safety culture in the maritime industry. *Safety, Reliability and Risk Analysis: Beyond the Horizon*, pp 467–473

- LaMontagne AD, Barbeau E, Youngstrom RA, Lewiton M, Stoddard AM, McLellan D, Wallace LM, Sorensen G (2004) Assessing and intervening on OSH programmes: effectiveness evaluation of the Wellworks-2 intervention in 15 manufacturing worksites. *Occup Environ Med* 61:651–660. <https://doi.org/10.1136/oem.2003.011718>
- Lappalainen J (2016) Finnish maritime personnel's conceptions on safety management and safety culture. PhD Thesis, Annales Universitatis Turkuensis, Turku, Finland
- Lappalainen J, Vepsäläinen A, Salmi K, Tapaninen U (2011) Incident reporting in Finnish shipping companies. *WMU J Marit Aff* 10:167–181
- Lappalainen FJ, Kuronen J, Tapaninen U (2014) Evaluation of the Ism Code in the Finnish Shipping Companies. *J Marit Res* 9:23–32
- Le Coze J-C, Pettersen K, Reiman T (2014) The foundations of safety science. *Saf Sci* 67:1–5. <https://doi.org/10.1016/j.ssci.2014.03.002>
- Leveson N (2011) Engineering a safer world: systems thinking applied to safety. MIT Press, Cambridge
- Li Y, Guldenmund FW (2018) Safety management systems: a broad overview of the literature. *Saf Sci* 103: 94–123. <https://doi.org/10.1016/j.ssci.2017.11.016>
- Lofquist EA (2017) Jousting with dragons: a resilience engineering approach to managing SMS in the transport sector. Discussion Paper 2017-19. International Transport Forum and Organization for Economic Co-operation and Development
- Maurino D (2017) Why SMS: an introduction and overview of safety management systems. ITF Discussion Paper 2017-16, International Transport Forum and Organisation for Economic Co-operation and Development
- Mayer IS, van Daalen CE, Bots PWG (2004) Perspectives on policy analyses: a framework for understanding and design. *Int J Technol Policy Manag* 4:169–191. <https://doi.org/10.1504/IJTPM.2004.004819>
- Meyer T, Reniers G (2016) Engineering risk management. Walter de Gruyter Graduate, Berlin, Munich, Boston 339p
- Möller N, Hansson SO, Holmberg JE, Rollenhagen C (2018) Handbook of safety principles. Wiley Essentials in Operations Research and Management Science, John Wiley and Sons, Hoboken
- Oltedal HA (2010) The use of safety management systems within the Norwegian tanker industry: do they really improve safety? *Reliab Risk Saf Theory Appl* 1-3:2355–2362
- Pantouvakis A, Karakasnakis M (2016) An empirical assessment of ISM Code effectiveness on performance: the role of ISO certification. *Maritime Policy and Management*
- Pawson R, Tilley N (1997) An introduction to scientific realist evaluation. In: Chelmsky E, Shadish WR (eds) *Evaluation for the 21st century: a handbook*. Sage Publications, Inc, Thousand Oaks, pp 405–418
- Pitchforth J, Mengersen K (2013) A proposed validation framework for expert elicited Bayesian Networks. *Expert Syst Appl* 40:162–167. <https://doi.org/10.1016/j.eswa.2012.07.026>
- Praetorius G, Hollnagel E, Dahlman J (2015) Modelling vessel traffic service to understand resilience in everyday operations. *Reliab Eng Syst Saf Spec Iss Resilience Eng* 141:10–21. <https://doi.org/10.1016/j.res.2015.03.020>
- Qureshi ZH (2007) A review of accident modelling approaches for complex socio-technical systems. Proceedings of the 12th Australian Workshop on Safety Related Programmable Systems, Adelaide, Australia
- Rae A (2015) Open questions and closed minds: mapping the gaps and divisions in the safety body of knowledge. Australian Safety Critical Systems Conference, Brisbane
- Rae D (2018) Risk and safety indicators. In: Möller N, Hansson SO, Holmberg JE, Rollenhagen C (eds) *Handbook of safety principles*. Wiley Essentials in Operations Research and Management Science, John Wiley and Sons, Hoboken
- Rae A, Provan D (2018) Safety work versus the safety of work. *Saf Sci*. <https://doi.org/10.1016/j.ssci.2018.07.001>
- Rae A, Alexander R, McDermid J (2014) Fixing the cracks in the crystal ball: a maturity model for quantitative risk assessment. *Reliab Eng Syst Saf*, Special issue of selected articles from ESREL 2012 125, 67–81. <https://doi.org/10.1016/j.res.2013.09.008>
- Reiman T, Rollenhagen C (2011) Human and organizational biases affecting the management of safety. *Reliab Eng Syst Saf* 96:1263–1274. <https://doi.org/10.1016/j.res.2011.05.010>
- Reiman T, Rollenhagen C (2014) Does the concept of safety culture help or hinder systems thinking in safety? *Accid Anal Prev* 68:5–15. <https://doi.org/10.1016/j.aap.2013.10.033>
- Rosa EA (2003) The logical structure of the social amplification of risk framework (SARF): metatheoretical foundation and policy implications. In: Pidgeon N, Kaspersen RE, Slovic P (eds) *The social amplification of risk*. Cambridge University Press, Cambridge

- Schröder-Hinrichs JU (2010) Human and organizational factors in the maritime world – are we keeping up to speed? *WMU J Marit Aff* 9(1):1–3
- Størkersen KV, Antonsen S, Kongsvik T (2017) One size fits all? Safety management regulation of ship accidents and personal injuries. *J Risk Res* 20(9):1154–1172
- Teo EAL, Ling FYY (2006) Developing a model to measure the effectiveness of safety management systems of construction sites. *Build Environ* 41(11):1584–1592
- Thomas MJW (2012) A systematic review of the effectiveness of safety management systems. Australian Transport Safety Bureau, Report No. AR-2011-148, 46 pp
- Tzannatos E, Kokotos D (2009) Analysis of accidents in Greek shipping during the pre- and post-ISM period. *Mar Policy* 33:679–684
- Valdez Banda OA, Goerlandt F (2017) The design of VTS Finland Safety Intent Specification. Aalto University publication series SCIENCE + TECHNOLOGY 8/2017
- Valdez Banda OA, Goerlandt F (2018) A STAMP-based approach for designing maritime safety management systems. *Saf Sci* 109:109–129. <https://doi.org/10.1016/j.ssci.2018.05.003>
- Valdez Banda OA, Hänninen M, Lappalainen J, Kujala P, Goerlandt F (2016) A method for extracting key performance indicators from maritime safety management norms. *WMU J Marit Aff* 15:237–265. <https://doi.org/10.1007/s13437-015-0095-z>
- Yücel G, van Daalen E (2009) An objective-based perspective on assessment of model-supported policy processes. *J Artif Soc Soc Simul* 12(4):3

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Osiris A. Valdez Banda¹ · Floris Goerlandt² · Johanna Salokannel³ · Pieter H. A. J. M. van Gelder⁴

✉ Osiris A. Valdez Banda
osiris.valdez.banda@aalto.fi

¹ Department of Mechanical Engineering (Marine Technology), Research Group on Maritime Risk and Safety, Aalto University, FI-00076 Aalto, Finland

² Department of Industrial Engineering, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

³ NOVIA University of Applied Science, Turku, Finland

⁴ Faculty of Technology, Policy and Management, Safety and Security Science Group, TU Delft, Delft, the Netherlands